

# 时序数据的异常检测可视化综述

## 1 介绍

时序数据被定义为一系列基于一个准确时间测量的结果，时间间隔通常是规律的[1]。[这里应该有几个数据集之类的说明下哪些数据有时序数据](#)

对于时序数据的分析在今天越来越广泛的应用在科学，工程，和商业领域，可视化帮助人们利用感知减少认知负荷进而理解数据[2]。长期以来，可视化也已经成功的被应用在对于时序数据的分析中来[3]。例如社交媒体[4]，城市数据[5]，电子交易[6]，时序排名[7]。在不同领域的时序数据中发现重要的特征和趋势的日益增长的需求刺激了许多可视交互探索工具的发展[8]：Line Graph Explore[9]，LiveRAC[2]，SignalLens[10]和 Data Vases[11]等。

时序数据的可视分析任务中，包括特征提取[14]，相关性分析和聚类[7]，模式识别[9]，异常检测[10]等。而异常检测在不同的研究领域都是一个重要的问题，异常检测表示发现数据中不符合预期行为的模式[12]。异常检测的目的是找到某些观察结果，它与其他观察结果有很大的偏差，以至于引起人们怀疑它是由不同的机制产生的[17]。对应到不同的领域中，网络安全中的异常表示网络设备异常或者可疑的网络状态[13]。情感分析中的异常表示一组数据中反常的观点，情绪模式，或者产生这些模式的特殊时间[16]。社交媒体中的异常可以是反常的行为，例如识别网络机器人[20]，反常的传播过程，例如谣言的传播[19]。这些异常信息或模式的产生原因，可能是会影响日常生活，社会稳定的因素，例如电脑侵入，社交机器人，道路拥堵状况等。提早发现识别这些异常有助于及时找到产生原因和实际状况，从而进一步分析或解决问题。

异常检测已经有许多成熟的方法，而且在机器学习领域也引起了广泛的关注[12]，包括有监督[21]和无监督的异常检测方法[22]。但是异常检测的会有一些挑战。当涉及到人工标注数据的问题时，往往需要大量的数据，费事费力，难以获取，同时又十分依赖于主观认为的判断，这些极大地影响了最后的分析结果质量[20]。与此同时，如何在自然数据中定义其中正常或异常的行为也是十分困难的[23]。这时就需要加入人的判断，可视化在其中可以更好的帮助人来分析理解数据和其中的行为，模式等，可以说，可视化在异常检测方面有得天独厚的优势。

2 挑战

异常检测的挑战在 Chandola V. [12]等人的综述中，被总结的很全面。

- a. 如何定义异常。正常和异常的界限往往是难以区分的，特别是当界限被规定后，在界限附近的异常观测，很容易把正常当做异常，亦或是把正常当做异常，例如一些设定阈值的异常检测方法，很难处理此种问题，需要其他的信息进行辅助判断。
- b. 有些异常的行为通常是人为的恶意操控，会模仿真实的行为，让异常的现象观测起来和正常现象一样，让异常检测的任务变得十分困难。  
例
- c. 随着发展和进步，许多领域的正常行为也在与时俱进，其概念可能在未来会失效。而且很多数据集都是复杂和动态的，例如社交媒体数据，网络安全数据等，这些挑战在[19]中都有提到。而在复杂多变时效性很高的场景中，需要人的监督和判断来进行异常的检测，分析，理解。
- d. 领域间的技术很难被应用于另一个领域，不同领域间的实际情况不一样，有些异常的产生在其它领域可能是正常的情况。
- e. 用于训练确认异常的模型的标记数据的获取以及可用性是很大的问题。
- f. 通常数据中包含的噪音和异常往往很相似，如何去区分和去除也是面临的问题。

数据

2.1 数据种类

2.2 数据属性

3 异常定义

4 异常检测方法(方法-可视化工具-数据源)

异常检测方法	领域	工作	数据源		可视化方法
聚类	社交媒体	[15]	Twitter		地理地图，标签云
对比	社交媒体	[18]	Twitter	对比信息内容和频率	

				和历史对比	
无监督机器学习	社交媒体	[19]	Twitter Facebook		
Self Organizing Maps (SOM),					
对比	网络攻击	[24]	the Darpa 1999		3D 图
阈值,训练模型		[25]			3D, 地理地图, 柱状图, 散点图, 平行坐标图
对比, 无监督学习		[20]	Twitter		高位投影, Glyph, 拓扑图, 热力图

## 5 数据处理

## 6 可视化任务, 应用和流程

### 6.1 可视化任务 (任务-可视化工具-特征-数据集)

### 6.2 可视化应用

### 6.3 可视化流程

## 7 时序数据异常检测可视化

[19]提供了多种视觉表现和视角用于对于异常转发现成的交互式探索系统, 支持对社交会话中异常的检测, 探索 and 解释, 它的分析模块结合了多种机器学习算法组来探索异常转发的特征, 并通过 OCCRF 模型来执行异常检测。此项工作基于 Facebook 和 Twitter 数据。

## 8 结论和未来工作

## 9 参考文献

- [1] Bojan V C, Raducu I G, Pop F, et al. Cloud-based service for time series analysis and visualisation in Farm Management System[C]//Intelligent Computer

- Communication and Processing (ICCP), 2015 IEEE International Conference on. IEEE, 2015: 425-432.
- [2] McLachlan P, Munzner T, Koutsofios E, et al. LiveRAC: interactive visual exploration of system management time-series data[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2008: 1483-1492.
  - [3] Weber M, Alexa M, Müller W. Visualizing time-series on spirals[C]//Infovis. 2001, 1: 7-14.
  - [4] Kumar P, Sinha A. Real-time analysis and visualization of online social media dynamics[C]//Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on. IEEE, 2016: 362-367.
  - [5] Chen W, Huang Z, Wu F, et al. VAUD: A Visual Analysis Approach for Exploring Spatio-Temporal Urban Data[J]. IEEE Transactions on Visualization & Computer Graphics, 2017 (1): 1-1.
  - [6] Xie C, Chen W, Huang X, et al. Vaet: A visual analytics approach for e-transactions time-series[J]. IEEE transactions on visualization and computer graphics, 2014, 20(12): 1743-1752.
  - [7] Xia J, Hou Y, Chen Y V, et al. Visualizing Rank Time Series of Wikipedia Top-Viewed Pages[J]. IEEE Computer Graphics and Applications, 2017, 37(2): 42-53.
  - [8] Cho M, Kim B, Bae H J, et al. Stroscope: Multi-scale visualization of irregularly measured time-series data[J]. IEEE transactions on visualization and computer graphics, 2014, 20(5): 808-821.
  - [9] Kincaid R, Lam H. Line graph explorer: scalable display of line graphs using focus+ context[C]//Proceedings of the working conference on Advanced visual interfaces. ACM, 2006: 404-411.
  - [10] Kincaid R. Signallens: Focus+ context applied to electronic time series[J]. IEEE Transactions on Visualization and Computer Graphics, 2010, 16(6): 900-907.
  - [11] Thakur S, Rhyne T M. Data vases: 2d and 3d plots for visualizing multiple time series[J]. Advances in Visual Computing, 2009: 929-938.
  - [12] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. ACM computing surveys (CSUR), 2009, 41(3): 15.

- [13] Pearlman J, Rheingans P. Visualizing network security events using compound glyphs from a service-oriented perspective[M]//VizSEC 2007. Springer, Berlin, Heidelberg, 2008: 131-146.
- [14] Alonso O, Khandelwal K. Kondenser: Exploration and visualization of archived social media[C]//Data Engineering (ICDE), 2014 IEEE 30th International Conference on. IEEE, 2014: 1202-1205.
- [15] Thom D, Bosch H, Koch S, et al. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages[C]//Pacific visualization symposium (PacificVis), 2012 IEEE. IEEE, 2012: 41-48.
- [16] Wang Z, Joo V, Tong C, et al. Anomaly Detection through Enhanced Sentiment Analysis on Social Media Data[C]//Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on. IEEE, 2014: 917-922.
- [17] Hawkins D M. Identification of outliers[M]. London: Chapman and Hall, 1980.
- [18] Schreck T, Keim D. Visual analysis of social media data[J]. Computer, 2013, 46(5): 68-75.
- [19] Zhao J, Cao N, Wen Z, et al. # FluxFlow: Visual analysis of anomalous information spreading on social media[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1773-1782.
- [20] Cao N, Shi C, Lin S, et al. Targetvue: Visual analysis of anomalous user behaviors in online communication systems[J]. IEEE transactions on visualization and computer graphics, 2016, 22(1): 280-289.
- [21] Steinwart I, Hush D, Scovel C. A classification framework for anomaly detection[J]. Journal of Machine Learning Research, 2005, 6(Feb): 211-232.
- [22] Eskin E, Arnold A, Prerau M, et al. A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data[J]. Applications of data mining in computer security, 2002, 6: 77-102.
- [23] Cao N, Lin Y R, Gotz D, et al. Z-Glyph: Visualizing outliers in multivariate data[J]. Information Visualization, 2017: 1473871616686635.
- [24] Onut I V, Zhu B, Ghorbani A A. A novel visualization technique for network anomaly detection[C]//PST. 2004: 167-174.
- [25] Riveiro M, Falkman G, Ziemke T. Improving maritime anomaly detection and situation awareness through interactive visualization[C]//Information Fusion, 2008 11th International Conference on. IEEE, 2008: 1-8.

